

## Clinical Focus

# Interactive Book Reading to Accelerate Word Learning by Kindergarten Children With Specific Language Impairment: Identifying Adequate Progress and Successful Learning Patterns

Holly L. Storkel,<sup>a</sup> Rouzana Komesidou,<sup>a</sup> Kandace K. Fleming,<sup>a</sup> and Rebecca Swinburne Romine<sup>a</sup>

**Purpose:** The goal of this study was to provide guidance to clinicians on early benchmarks of successful word learning in an interactive book reading treatment and to examine how encoding and memory evolution during treatment contribute to word learning outcomes by kindergarten children with specific language impairment (SLI).

**Method:** Twenty-seven kindergarten children with SLI participated in a preliminary clinical trial using interactive book reading to teach 30 new words. Word learning was assessed at 4 points during treatment through a picture naming test.

**Results:** The results indicate that the following performance during treatment was cause for concern, indicating a need to

modify the treatment: naming 0–1 treated words correctly at Naming Test 1; naming 0–2 treated words correctly at Naming Test 2; naming 0–3 treated words correctly at Naming Test 3. In addition, the results showed that encoding was the primary limiting factor in word learning, but memory evolution also contributed (albeit to a lesser degree) to word learning success.

**Conclusion:** Case illustrations demonstrate how a clinician's understanding of a child's word learning strengths and weaknesses develop over the course of treatment, substantiating the importance of regular data collection and clinical decision-making to ensure the best possible outcomes for each individual child.

Evidence-based practice seeks to integrate research evidence, clinician experience, and family and client perspectives into service delivery (American Speech-Language-Hearing Association [ASHA], n.d.). That is, clinicians are to select high-quality treatments based on research evidence and their own expertise. In addition, the clinician communicates the different treatment options to his or her client and family to gain an understanding of the client's and family's values and needs. This information is integrated to select a treatment approach. The clinician then monitors the effectiveness of the treatment as it is implemented to ensure that the predicted effectiveness of the treatment is actually occurring. This monitoring function has the potential to create challenges for clinicians. To be specific, although collecting data during treatment may be relatively easy, evaluating progress for clinical

decision making is potentially more difficult, requiring a clinician to have in mind benchmarks for success for clients from the target population at different points during treatment. For this article, we take up the case of targeting a word learning deficit for a kindergarten child with specific language impairment (SLI) and walk through the process of selecting the treatment, establishing benchmarks for adequate progress, determining treatment modifications, and then integrating this information for clinical decision making.

## Selecting the Treatment

Children with SLI are slow to learn new words, needing 2–3 times as many exposures as their peers (Gray, 2003; Rice, Oetting, Marquis, Bode, & Pae, 1994). This difficulty with word learning leads to deficits in both breadth (number of words known) and depth (detailed knowledge of words) of vocabulary throughout the school-age years (McGregor, Oleson, Bahnsen, & Duff, 2013). Vocabulary deficits, in turn, affect reading decoding and comprehension (Ouellette, 2006), leading these children to fall further behind the academic achievement of their peers (Morgan, Farkas, & Wu, 2011). In addition to this academic cost,

<sup>a</sup>University of Kansas, Lawrence

Correspondence to Holly L. Storkel: hstorkel@ku.edu

Editor: Shelley Gray

Associate Editor: Monique Mills

Received July 18, 2016

Revision received October 31, 2016

Accepted December 5, 2016

[https://doi.org/10.1044/2017\\_LSHSS-16-0058](https://doi.org/10.1044/2017_LSHSS-16-0058)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

there is a social cost to vocabulary deficits, with low vocabulary being linked to low popularity among peers (Gertner, Rice, & Hadley, 1994). Despite the significant cost of vocabulary deficits, there are few treatments that have been proven effective for children with SLI (Cirrin & Gillam, 2008).

One effective treatment for word learning generally is *interactive book reading*. Interactive book reading involves an adult reading a book to a child and deviating from the text to provide explicit instruction (e.g., define the new word). It is important to note that randomized clinical trials, meta-analyses, and systematic reviews show that interactive book reading has moderate-to-large effects on word learning by typically developing children and children with low vocabulary due to environmental differences in input (i.e., children from low-income families; Justice, Meier, & Walpole, 2005; Marulis & Neuman, 2010; Mol, Bus, & de Jong, 2009; Mol, Bus, de Jong, & Smeets, 2008; Whitehurst et al., 1988). However, results from children with different language impairments have been less robust (Crain-Thoreson & Dale, 1999; Dale, Crain-Thoreson, Notari-Syverson, & Cole, 1996; Pile, Girolametto, Johnson, Chen, & Cleave, 2010; Whitehurst et al., 1991), although at least one study shows that parents can facilitate word use by children with language impairments (Crowe, Norris, & Hoffman, 2004).

In a recently completed preliminary clinical trial (Storkel et al., 2017), we proposed that a crucial piece of information was missing when applying interactive book reading to children with SLI: the appropriate intensity of the treatment. Thus, an interactive book reading treatment that was proven effective for kindergarten children from low-income families (Justice et al., 2005) was tested at different intensities with kindergarten children with SLI to determine which intensity was the most promising. The specific intensities tested were 12 (which was the intensity that was effective in the prior study with low-income children), 24, 36, and 48 exposures to the treated words. A definition task was used to determine which treated words were learned. Results showed that response to treatment improved as intensity increased from 12 to 24 to 36 exposures, and then no further improvements were observed as intensity increased to 48 exposures. Thus, 36 exposures was judged to be the most promising of the four tested intensities. In terms of specific outcomes in the 36-exposure condition, 43% of children showed a positive response to treatment (operationalized as correctly defining five or more words), and children in this condition, on average, correctly defined five treated words. This compares somewhat favorably to the prior study by Justice et al., where 72% of children from low-income families showed a positive response to treatment (operationalized as a 4-point gain on a definition task), and children gained, on average, approximately six words. Although this is a promising start, clearly further development of this treatment approach is needed to expand the number of children with SLI who respond positively to the treatment and to further increase the number of new words learned.

Although the evidence to support the use of interactive book reading in treating word learning deficits by

children with SLI certainly is preliminary, this case illustrates a typical clinical conundrum. The evidence base lags behind our clinical needs. Yet, clinicians cannot place clients on hold until a well-supported treatment is identified. Thus, preliminary treatments may need to be used in clinical practice. This is appropriate provided that there is evidence the client is making adequate progress with the treatment, but we must determine what constitutes adequate progress. In this article, we use data from our preliminary clinical trial to identify benchmarks for adequate treatment progress (Storkel et al., 2017) to demonstrate what researchers can do to provide clinicians with useful information from preliminary clinical research and to illustrate options that clinicians could use to collect their own local data for benchmarking progress.

### ***Determining Benchmarks for Adequate Progress***

To determine benchmarks, we first classify the 27 children in our preliminary clinical trial based on their treatment outcomes and then examine their earlier progress during treatment.

### **Participants**

Twenty-seven kindergarten children with SLI (age:  $M = 5;8$  [months;years],  $SD = 0;6$ , range 5;0–6;5; 52% girls) participated in the preliminary clinical trial to determine the adequate intensity of interactive book reading to teach new words to children with SLI. Pre- and post-treatment definition data for these same children were reported in Storkel et al. (2017), with the goal of that report being to establish the adequate intensity of interactive book reading for children with SLI. The current report focuses on picture naming data taken during treatment and at the end of treatment for the same children. Children were recruited through school-based language screenings conducted by the research team (52%), referral by a speech-language pathologist or kindergarten teacher (41%), or by public announcement (7%). Children were required to (a) be enrolled in kindergarten or eligible for kindergarten; (b) pass a hearing screening (ASHA, 1997); (c) score at or above the 16th percentile for nonverbal cognition as measured by the Reynolds Intellectual Assessment Scale (Reynolds & Kamphaus, 2003); (d) have a Core Language Score at or below the 10th percentile on the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003); and (e) score at or below the 10th percentile on at least one of three vocabulary measures. The 10th percentile (approximately 1.25  $SDs$  below the mean) was chosen for the CELF-4 Core Language because adequate sensitivity and specificity ( $>0.80$ ) are shown for scores near this cutoff (i.e., 1–1.5  $SDs$  below the mean), according to the test manual. The same criterion was used for the vocabulary measures even though test manuals did not report sensitivity and specificity for these specific measures.

Table 1 shows the participant characteristics. Note that the CELF-4 Core Language Score is heavily weighted towards expressive language (i.e., three of the four subtests

**Table 1.** Percentile scores for participants on standardized clinical tests.

Test	Mean	SD	Range	% at or below 10th percentile
RIAS <sup>a</sup> Nonverbal IQ	55	25	23–99	0
CELF <sup>b</sup> Core Language	3	3	0.1–10	100
Vocabulary: DELV <sup>c</sup> Semantic	9	8	0.1–25	74
Vocabulary: CELF <sup>b</sup> Word Classes	24	19	1–75	37
Vocabulary: CREVT <sup>d</sup> Expressive	26	16	1–63	19
CELF <sup>b</sup> Understanding Spoken Paragraphs	7	8	0.1–25	81
CTOPP <sup>e</sup> Nonword Repetition	22	21	1–75	44
CTOPP <sup>e</sup> Phonological Memory	15	17	1–75	52
CTOPP <sup>e</sup> Phonological Awareness	8	8	1–30	74
GFTA <sup>f</sup>	30	21	1–67	19

<sup>a</sup>Reynolds Intellectual Assessment Scale (Reynolds & Kamphaus, 2003); <sup>b</sup>Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003); <sup>c</sup>Diagnostic Evaluation of Language Variation (Seymour, Roeper, de Villers, & de Villers, 2005); <sup>d</sup>Comprehensive Receptive and Expressive Vocabulary Test–Third Edition (Wallace & Hammill, 2013); <sup>e</sup>Comprehensive Test of Phonological Processing–Second Edition (Wagner, Torgesen, Rashotte, & Pearson, 2013); <sup>f</sup>Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000).

used to compute the score are expressive tests). Indeed, all of the participants scored below the 10th percentile on the CELF-4 Expressive Language Index, indicating that all children had expressive language deficits. As shown in Table 1, the majority (81%) also had deficits in aspects of receptive language as measured by the CELF-4 Understanding Spoken Paragraphs subtest. This subtest was used because it required many of the same skills as interactive book reading. Many children demonstrated deficits (i.e., performance below the 10th percentile) in other aspects of language, including phonological awareness (74%), phonological memory (52%), nonword repetition (44%), and (to a lesser extent) articulation (19%).

Demographic characteristics generally matched those of the recruitment area (Eastern Kansas). The race and ethnicity of the children was 63% White-Non-Hispanic, 19% White-Hispanic, 11% Black/African American–Non-Hispanic, 4% White–Unknown Ethnicity, and 4% Unknown Race and Ethnicity. The marital status of parents was 70% married, 19% single, and 11% divorced. Mother’s education was 37% partial college, 30% college graduates, 22% high school graduates, 4% partial high school, 4% graduate degrees, and 4% unknown. Father’s education was 37% not reported (mostly from the single/divorced families), 22% high school graduates, 22% partial college, 7% college graduates, 4% junior high school, 4% partial high school, and 4% graduate degrees.

## Treatment

The treatment is described in greater detail in Storkel et al. (2017) and Voelmlé and Storkel (2015). Likewise, some of the treatment materials (i.e., treatment stimuli, treatment scripts, treatment schedules) are available in the KU ScholarWorks archive at <http://hdl.handle.net/1808/20313>. The main goal of the preliminary clinical trial was to determine the adequate intensity of interactive book reading for children with SLI to learn new words. Thus, children were randomly assigned to one of four possible treatment

intensities. Intensity was operationally defined as the total number of exposures to the treated words. The four intensities were 12, 24, 36, or 48 cumulative exposures. Within a given intensity (12, 24, 36, or 48 cumulative exposures), children were randomly assigned to one of two possible sets of five commercially available storybooks. These storybooks were used in a prior clinical trial of interactive book reading for children at risk for vocabulary deficits due to low income (Justice et al., 2005). The prior study identified six words in each book that were unlikely to be known by kindergarten children (Justice et al., 2005). Pretreatment testing in that study showed that only 10% or fewer of children provided at least a partially accurate definition for 80% of the words (24 of 30). Within a set of five treated books, there were 30 words that served as the focus of treatment. The set of books and words that was not chosen for treatment was repeatedly tested and served as untreated control words. See the Summary of the Stimuli file in the KU ScholarWorks archive for the list of books and target words.

Children participated in two treatment sessions per week, with the total number of treatment sessions varying from 10 to 20 (approximately 5–10 weeks) depending on the assigned intensity (12, 24, 36, or 48 cumulative exposures). Treatment sessions lasted approximately 20–30 min and occurred in the child’s school, home, or other agreed upon location (e.g., local library). Treatment was provided by a paid research assistant, typically a PhD student. Each treatment session targeted two storybooks, with one storybook being the focus of the first half of the session and the second storybook being the focus of the second half of the treatment session. Treatment for each book began with a preview activity. During preview, a color picture depicting each treated word was shown individually while the researcher read from a script. The exact exposure during preview varied by intensity but typically included a definition (e.g., “*Marsh* means a low, wet land, often thick with tall grasses”) and a synonym (“*Marsh* is like a swamp”). After all treated words were previewed, the storybook was read.

In some intensities, no additional teaching occurred during book reading, whereas in higher intensity conditions the synonym of the target word was provided again (e.g., reading, “They came down to a *marsh*, where they saw a muskrat spring cleaning his house”; additional input, “*Marsh* is like a swamp”). Upon completing book reading, a review activity was initiated. A different color picture depicting each treated word was shown individually while the researcher read from the script. Again, the exact exposure during review varied by intensity but always included a context sentence that matched the review picture (e.g., “Ducks and beavers live in a *marsh* because they like the water.”) and in higher intensities also included a reminder of the definition (e.g., “*Marsh* means a low, wet land, often thick with tall grasses.”).

In a session, children heard each target word three to six times depending on the intensity (see the Treatment and Naming Scripts file in the KU ScholarWorks archive). In addition, each book was the focus of four to eight sessions depending on the intensity (see the Treatment Schedules file in the KU ScholarWorks archive).

### Outcome Measures

Both a definition and a naming task were used to examine learning of treated and untreated control words. The focus for this study is the naming task because it was administered during treatment by the treatment provider to monitor progress, whereas the definition task was administered only before and after treatment (see Storkel et al., 2017, for a detailed report of the definition data). The naming task was administered at four predetermined points during treatment, as shown in Table 2. The naming task was administered during treatment sessions. The naming task tested the words that were the focus of treatment for the given session as well as a paired set of untreated control words. This pairing was accomplished by randomly pairing an untreated book with a treated book. Children were shown the preview picture used in treatment and were given a prompt specific to the picture and the target word. For example, for the target word *ruffle*, children were shown a picture of a bird and asked, “What does the bird do to his feathers?” Naming prompts are shown in the KU ScholarWorks archive (see Treatment and Naming Scripts

**Table 2.** For each intensity, the schedule for administering the naming task in terms of number of exposures and repeated book reading.

Intensity	Test 1	Test 2	Test 3	Test 4
12				
Exposures	3	6	9	12
Reading	First	Second	Third	Fourth/last
24				
Exposures	4	12	20	24
Reading	First	Third	Fifth	Sixth/last
36				
Exposures	6	12	24	36
Reading	First	Second	Fourth	Sixth/last
48				
Exposures	12	24	36	48
Reading	Second	Fourth	Sixth	Eighth/last

file). Responses were transcribed and scored as correct if the child named the target word (e.g., said “ruffle” for target *ruffle*) or incorrect if the child failed to name the target word (e.g., said “bird” for target *ruffle*) or failed to provide any response. Changes in grammatical form (e.g., “ruffled” or “ruffling” for target *ruffle*) and common misarticulations (e.g., “wuffle” for target *ruffle*) were scored as correct.

### Overall Treatment Response Classification

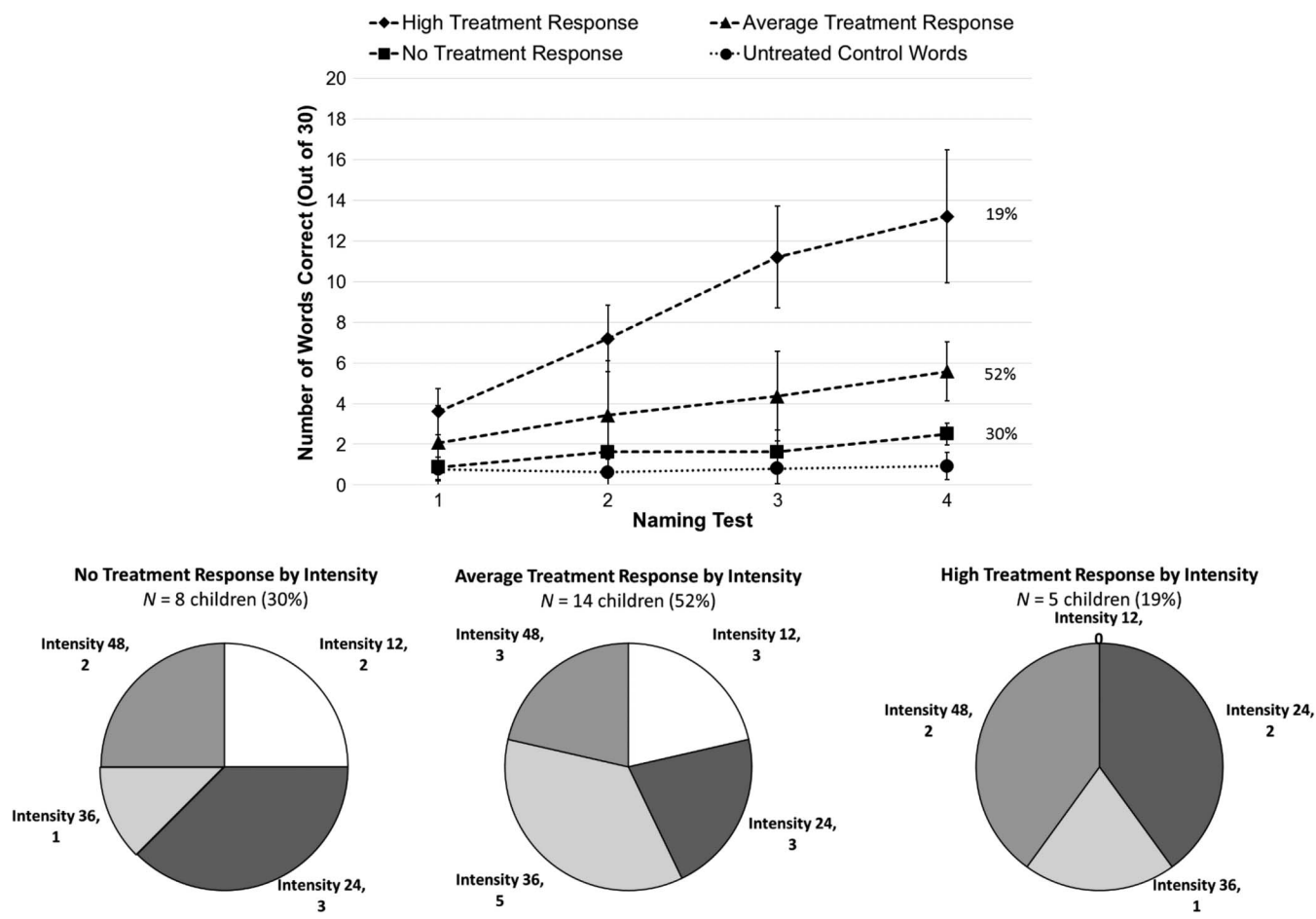
The first step in identifying benchmarks for adequate progress is to classify each child’s treatment response at the end of treatment. The number of treated words correct at the last naming test was used to classify the child’s overall response to the treatment as (a) no response, (b) average response, and (c) high response. Group performance on untreated and treated words was used to set an operational definition for each treatment response category. To be specific, children who named zero to three words correctly at the last naming test were classified as having no response to treatment because this performance was within the range observed for untreated control words ( $M = 1$ ,  $SD = 1$ , range 0–3). Children who named four to nine words correctly at the last naming test were classified as average responders because this performance was outside the range observed for untreated control words ( $M = 1$ ,  $SD = 1$ , range 0–3) but within 1  $SD$  of the mean number of treated words named correctly ( $M = 6$ ,  $SD = 4$ , range 2–18). Last, children who named 10 or more words correctly at the last naming test were classified as high responders because this performance was 1  $SD$  (or more) above the mean number of treated words named correctly.

Figure 1 shows data from the four naming tests for the untreated control words for all children (short-dashed line). As expected, children with SLI named few untreated words correctly, indicating that they were not learning these words over time on their own. Eight of the 27 children (30%) were classified as nonresponders. As shown in Figure 1, these children’s naming of treated words at the end of treatment ( $M = 2.5$ ,  $SD = 0.5$ , range 2–3) was similar to naming of untreated words ( $M = 0.9$ ,  $SD = 0.7$ , range 0–3), indicating that the treatment was not successfully supporting learning for these children. In contrast, 14 of the 27 children (52%) were classified as average responders. Figure 1 shows that these children named a modest number of treated words correctly at the end of treatment ( $M = 5.6$ ,  $SD = 1.5$ , range 4–8) and that this learning was higher than that observed for untreated words. This pattern indicates that the treatment was supporting at least modest learning. Last, 5 of the 27 children (19%) were classified as high responders. Figure 1 shows that these children named many treated words correctly at the end of treatment ( $M = 13.2$ ,  $SD = 3.3$ , range 10–18). This learning was much higher than that of untreated words, indicating stronger learning for these children.

The bottom panels of Figure 1 show the treatment intensity of the children within each treatment response classification. In general, a mix of treatment intensities are present within each treatment response group. Although



**Figure 1.** Top panel shows the average number of words correct, with error bars indicating the standard deviation for untreated control words for all children (circles, short-dashed line) and treated words (long-dashed line) for children with no treatment response (squares), an average treatment response (triangles), or a high treatment response (diamonds). Bottom panels show the percent of children in each intensity condition (12, 24, 36, or 48 exposures) for each treatment response group: no treatment response (left), average treatment response (middle), and high treatment response (right).



36 exposures was deemed the most promising intensity for children with SLI as a group, Figure 1 shows that there is individual variation in the intensity needed by each child. Thus, 36 exposures is a good starting intensity for interactive book reading, but client progress clearly should be used to verify whether that intensity is supporting adequate progress.

#### Identifying Adequate Treatment Progress Early in Treatment

Now that the treatment response for Naming Test 4 at the end of treatment has been classified, the next step is to examine earlier progress at Naming Tests 1, 2, and 3. A classification function was used to identify cutoffs for adequate progress at Naming Tests 1, 2, and 3. Classification functions are typically used to evaluate diagnostic measures. In the diagnostic case, the diagnostic status of participants is known (i.e., SLI vs. normal language), and then score(s) on a new diagnostic measure are used to predict the child's language status, based on a selected cutoff

score. The alignment between the known and predicted disorder status is then examined. The logic when applied to treatment progress is similar and is illustrated in Table 3. As noted previously, data from Naming Test 4 were used to classify each child's response to treatment. For the classification function, we differentiate the eight children who did not respond to treatment (30%) from the 19 children who did respond to treatment (70%), regardless of whether the response was average or high. Thus, the treatment outcome is known (see columns 2 and 3 of Table 3). At each of the earlier naming tests during treatment (i.e., Naming Tests 1, 2, and 3), a cutoff score can be selected. Children scoring at or above the cutoff score are predicted to respond to the treatment, whereas children scoring below the cutoff score are predicted to not respond to the treatment. This predicted outcome can then be compared with the known outcome, as shown in Table 3.

In some cases, the prediction will be incorrect. There are two ways that a prediction can be incorrect. On the

**Table 3.** Classification table and measures for analysis.

Early naming test prediction	Treatment outcome on last naming test	
	Did not respond to treatment (-): (3 or fewer words correct)	Did respond to treatment (+): (4 or more words correct)
At or above the criterion score = Predicted to respond to treatment (+) <i>Consequence</i>	False positive <i>Fail to modify treatment when modification is necessary.</i>	True positive <i>Correctly continue treatment without modification.</i>
Below the criterion score = Predicted to not respond to treatment (-) <i>Consequence</i>	True negative <i>Correctly modify treatment to attempt to improve outcome.</i>	False negative <i>Modify treatment when modification isn't necessary.</i>

**Measures for Analysis.** (1) Accuracy = (true positive + true negative) / total. (2) Sensitivity = true positive / (true positive + false negative), probability that the early naming measure predicts a child will respond to the treatment when the child actually did respond to treatment. (3) Specificity = true negative / (true negative + false positive), probability that the naming measure predicts a child will not respond to the treatment when the child actually does not respond to treatment. (4) Positive likelihood ratio = true positive / false positive, indicates how likely a child is to respond to treatment as opposed to not respond to treatment given an early naming score at or above the cutoff. (5) Negative likelihood ratio = false negative / true negative, indicates how likely a child is to respond to treatment rather than not respond to treatment given an early naming score below the cutoff. (6) Positive predictive value = true positive / (true positive + false positive), probability that the child will respond to treatment when an early naming score is at or above the cutoff. (7) Negative predictive value = true negative / (true negative + false negative), probability that the child will not respond to treatment when an early naming score is below the cutoff.

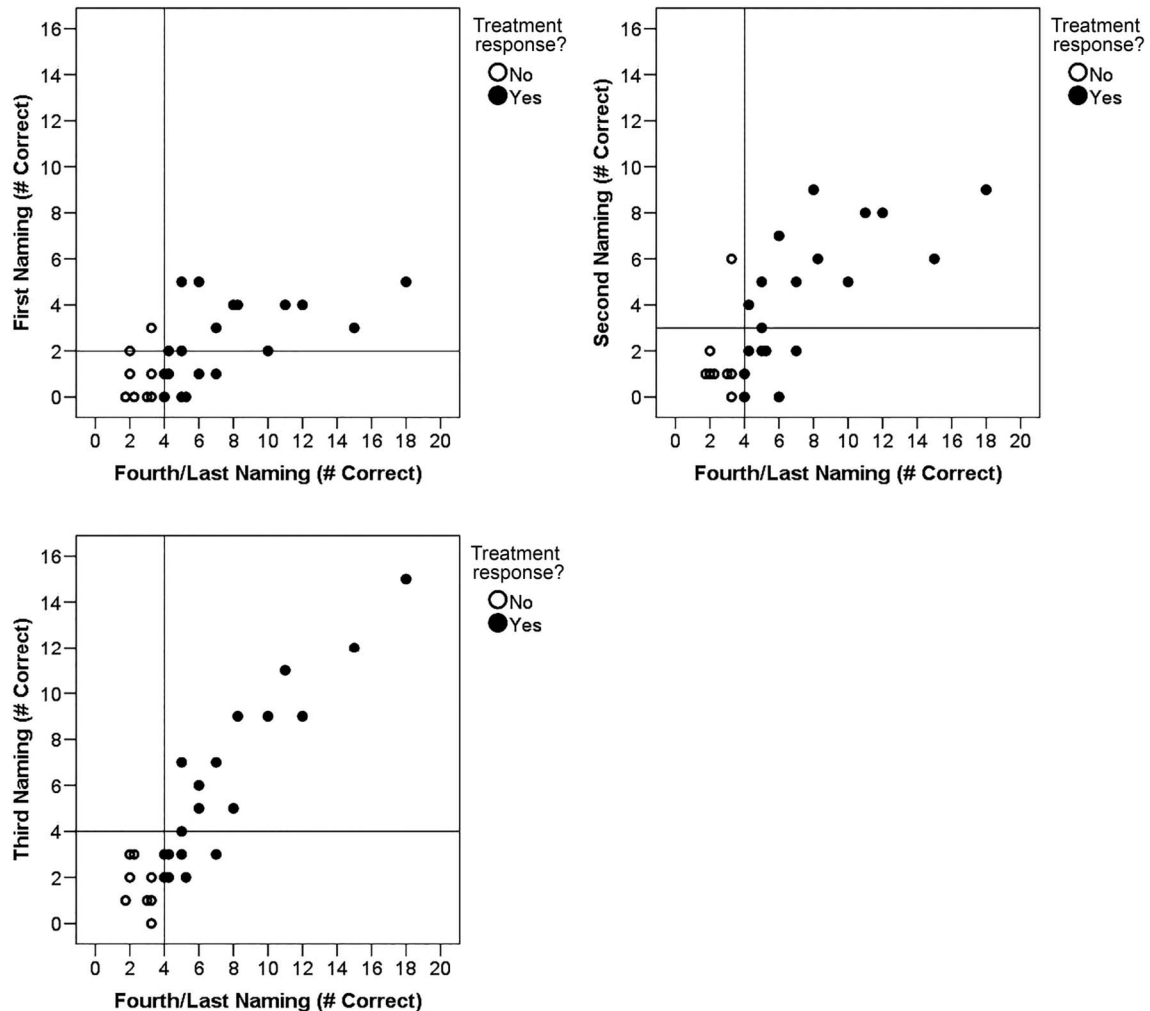
**Note.** See [https://www.medcalc.org/calc/diagnostic\\_test.php](https://www.medcalc.org/calc/diagnostic_test.php) for a program to calculate these measures.

one hand, one might predict that a child *will respond* to treatment when the child actually *will not respond* to treatment, namely a *false positive*. On the other hand, one might predict that a child *will not respond* to treatment when the child actually *will respond* to treatment, namely a *false negative*. It is important to weigh the cost of these inaccurate predictions to determine how to best select a cutoff score. That is, raising or lowering the cutoff score will increase one of these errors while minimizing the other. In selecting cutoff scores for Naming Tests 1, 2, and 3, false positives were minimized because these errors were judged to be costlier than false negatives. The rationale is that modifying treatment may not increase the cost of the treatment in terms of time spent in treatment—that is, in many cases, a clinician may spend the same amount of time in treatment with the child but would choose to do different treatment activities during that time to better support the child's learning. In this way, modifying treatment when modification isn't really necessary, as would be the case for a false negative, is not viewed as a highly costly mistake. In contrast, failing to modify a treatment when modification is needed, as would occur for a false positive, is viewed as a highly costly mistake because both the child and clinician would spend many hours in treatment but ultimately would not achieve the desired outcome. Thus, all the time invested in treatment would produce negligible results. Because the selected cutoff scores minimize false positives, classification measures related to false positives (i.e., specificity, positive likelihood ratio, positive predictive value) will tend to be in a more desirable range than those related to false negatives (i.e., sensitivity, negative likelihood ratio, negative predictive value), as described in Table 3.

Figure 2 shows the classification of children who did not respond to treatment (open circles) and children who did respond to treatment (filled circles) for each of the three

naming tests, and Table 4 summarizes the classification measures for each of the three naming tests. In general, accuracy in predicting treatment outcome improved as treatment progressed. For the first naming test, two words correct was selected as the cutoff. Thus, children who correctly named two or more words at the first naming test were predicted to respond to treatment, whereas children who correctly named zero to one words at the first naming test were predicted to not respond to treatment. In clinical practice, a clinician could use this benchmark to differentiate children who are making adequate progress in this particular treatment (i.e., those predicted to respond to treatment) from children who are not making adequate progress in this particular treatment (i.e., those predicted to not respond to treatment). As shown in the upper left panel of Figure 2, this cutoff score correctly predicted a treatment response for 12 children (i.e., true positive, data points in upper right of the graph) and correctly predicted a non-response for six children (i.e., true negative, data points in lower left of the graph). However, two children who did not respond to treatment were incorrectly predicted to respond to treatment (false positive, data points in upper left of the graph), and seven children who did respond to treatment were incorrectly predicted to not respond to treatment (false negative, data points in lower right of the graph). For each subsequent naming test, the number of false positives decreased by one child (i.e., one false positive for Naming Test 2, zero false positives for Naming Test 3), whereas the number of false negatives remained constant at seven children. Overall, the probability that an early naming score predicts a child will not respond to treatment when the child actually does not respond to treatment (i.e., specificity) was relatively high, indicating that these cutoff scores can potentially be used to guide clinical decision making when implementing our version of interactive book reading with

**Figure 2.** Fourth/last naming test scores for treated words plotted by first (upper left panel), second (upper right panel), and third (lower left panel) naming test scores. Vertical line indicates the criteria for a treatment response based on the fourth/last naming test, which is also shown by the color of the circles, with unfilled corresponding to no treatment response and filled corresponding to a treatment response. The horizontal line indicates the cutoff score on the earlier naming test for predicting a positive or negative response to treatment. Within a panel, the cutoff lines divide the data into four quadrants representing false positives (upper left), true positives (upper right), true negatives (lower left), and false negatives (lower right).



**Table 4.** Classification measures for absence versus presence of a treatment response ( $n = 27$  children).

Naming test	Cutoff <sup>a</sup> (No. of words correct)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Positive likelihood ratio	Negative likelihood ratio	Positive predictive value (%)	Negative predictive value (%)
1	2	67	63	75	2.53	0.49	86	46
2	3	70	63	88	5.05	0.42	92	50
3	4	74	63	100	N/A	0.37	100	53

*Note.* Correlation between Naming Test 1 and 4 (i.e., treatment response):  $r(27) = .62$ ,  $p = .001$ ,  $r^2 = .29$ ; correlation between Naming Test 2 and 4 (i.e., treatment response):  $r(27) = .74$ ,  $p < .001$ ,  $r^2 = .54$ ; correlation between Naming Test 3 and 4 (i.e., treatment response):  $r(27) = .93$ ,  $p < .001$ ,  $r^2 = .86$ .

<sup>a</sup>Cutoff refers to the criterion score used to predict the presence versus absence of a response to treatment. Scores at or above the cutoff predict that the child will respond to treatment, whereas scores below the cutoff predict that the child will not respond to treatment.

kindergarten children with SLI. Thus, modifications to this specific treatment should be considered for children who correctly name only zero to one words at Naming Test 1, zero to two words at Naming Test 2, or zero to three words at Naming Test 3.

### **Treatment Modifications**

If it is determined that a child is not making adequate progress, then the clinician must consider the underlying cause for the lack of progress and adjust the treatment accordingly. In this case, one must consider where the breakdown in learning is occurring.

Learning can be decomposed into two processes. The first process is referred to as *encoding*, which occurs as a person is receiving input during training. For word learning, encoding requires that the learner extract the novel word form and meaning from ongoing speech, hold this information in working memory, and store an initial representation of the word form and meaning in memory (McGregor, Arbisi-Kelm, & Eden, 2016; McGregor, Licandro, et al., 2013). Encoding during training appears to be swift and robust in typical children and adults (cf. Davis, Di Betta, Macdonald, & Gaskell, 2009; Davis & Gaskell, 2009; Henderson, Weighall, & Gaskell, 2013; McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000; Storkel, 2015). However, children with SLI appear to struggle with learning from input, requiring greater input to achieve the same outcomes as their peers (Gray, 2003; McGregor et al., 2016; McGregor, Licandro, et al., 2013; Rice et al., 1994; Riches, Tomasello, & Conti-Ramsden, 2005). Moreover, McGregor and colleagues (McGregor et al., 2016; McGregor, Oleson, et al., 2013) suggested that difficulty with encoding may be the main factor limiting word learning by adults with a history of language impairment, although this claim has yet to be extended to children with SLI (but see Alt, 2011, for claims about encoding difficulties in working memory for children with SLI).

The second process in learning is referred to as *memory evolution*, which occurs in the absence of input. Memory evolution refers to the mental processes that occur once a treatment session has ended and involves transferring the initial representation, which is thought to reside in the hippocampus, to the relevant language areas in the cortex during sleep (Walker & Stickgold, 2010). The new memory is then integrated with similar memories, specifically known words that share sound structure or meaning with the new word. In some cases, this transfer and integration can strengthen a new memory, leading to similar or even improved performance after the gap in training (i.e., retention; Gaskell & Dumay, 2003; Rice et al., 1994; Storkel, 2001, 2003; Storkel & Lee, 2011). In other cases, this transfer and integration can weaken a new memory, leading to poorer performance (i.e., forgetting; Storkel, Bontempo, Aschenbrenner, Maekawa, & Lee, 2013; Tamminen & Gaskell, 2013; Vlach & Sandhofer, 2012). This variability in the outcome of memory evolution indicates that it is a major point of vulnerability along the pathway to learning a new word (Storkel,

2015). Moreover, children with SLI appear to have difficulty retaining new learning (Adi-Japha & Abu-Asba, 2014; McGregor, Licandro, et al., 2013; Oetting, 1999; Rice et al., 1994; Riches et al., 2005). In fact, word learning by children with SLI appears to worsen as training accumulates, potentially due to deficits in memory evolution (Kan & Windsor, 2010).

At present, there is evidence that children with SLI potentially could struggle with encoding during treatment sessions and with retaining new words across treatment sessions (Adi-Japha & Abu-Asba, 2014; Gray, 2003; McGregor et al., 2016; McGregor, Licandro, et al., 2013; Oetting, 1999; Rice et al., 1994; Riches et al., 2005). Thus, if a child is not making adequate progress during interactive book reading treatment, the next logical question is whether the child is struggling with encoding, memory evolution, or both. Answering this question will lead the clinician to alter treatment in different ways to better support either encoding or memory evolution or both (Komesidou & Storkel, 2015). To address this issue, we use our naming data to classify children as having low, average, or high encoding as well as low, average, or high memory evolution based on their rank within the larger group. In addition, encoding and memory evolution are then examined for each of the treatment response groups to determine early indicators of success.

### **Encoding Scores**

Data from KAW018 (who was randomized to 24 cumulative exposures) are shown in Figure 3 and are used to illustrate the scoring procedures for encoding and memory evolution. Figure 3 shows the data from each of the four naming tests, with correct responses noted by a score of 1 and incorrect responses noted by a score of 0. Of the 30 treated words, only the words with at least one correct response are shown, excluding the treated words that were never named correctly. The encoding score was the number of words named correctly at least one time at any point during treatment. As shown in Figure 3, KAW018 named 12 words (i.e., *invisible*, *pouted*, *snuggled*, *squawked*, *flashing*, *haddock*, *smooth*, *tailor*, *worn*, *crept*, *gulp*, and *tight*) correctly at least once during naming testing. This score is thought to reflect encoding because the child has just completed the training for a given word. If the child has successfully encoded the word, he or she will name it correctly during the test. In contrast, if the child was not successful in encoding the word, he or she will not name it correctly.

### **Memory Evolution: Retention Scores**

The retention score examines whether a word continues to be named correctly after the first correct response. In Figure 3, this is indicated by shading. For example, the word *invisible* is named correctly at the first naming test. This first correct response would be tallied in the previously described encoding score. *Invisible* is named correctly again at the second and fourth/last naming tests. Thus, *invisible* is scored as a word that was retained. In contrast, the word *worn* is named correctly at the second test but is never named correctly again. Therefore, *worn* is scored as a word that



**Figure 3.** Partial naming data from the four test points for KAW018 (intensity: 24 cumulative exposures). 0 indicates an incorrect response. 1 indicates a correct response. A repeated correct response is noted by shading. Only a subset of the 30 treated words are shown, namely the treated words that were named correctly at least one time.

Treated Words	Naming Test 1: 4 Exposures	Naming Test 2: 12 Exposures	Naming Test 3: 20 Exposures	Naming Test 4: 24 Exposures
invisible	1	1	0	1
pouted	1	1	1	0
snuggled	1	1	1	1
squawked	1	1	0	1
flashing	0	1	1	1
haddock	0	1	0	1
smooth	0	1	0	1
tailor	0	1	0	1
worn	0	1	0	0
crept	0	0	1	0
gulp	0	0	1	0
tight	0	0	0	1

was not retained. Last, words that are named correctly for the first time at the last test have no opportunity to show retention because no further testing is available for those words, meaning that it is unknown whether those words would be named correctly again (i.e., retained). Therefore, these words are excluded from the retention score. To compute the retention score for KAW018, the number of words named correctly before the fourth/last test are counted, yielding 11 words (i.e., *invisible*, *pouted*, *snuggled*, *squawked*, *flashing*, *haddock*, *smooth*, *tailor*, *worn*, *crept*, and *gulp*, but not *tight*) that could have been retained or forgotten. The number of these words that were named correctly at a subsequent test is counted, yielding eight words (i.e., *invisible*, *pouted*, *snuggled*, *squawked*, *flashing*, *haddock*, *smooth*, and *tailor*, but not *worn*, *crept*, and *gulp*). Last, the percentage of words retained is computed, specifically  $8/11 = 73\%$ . Note that this is not a pure measure of retention because the child has received additional treatment between the first correct response and the subsequent correct response, providing an opportunity to re-encode the word. Therefore, although we term this score retention, it actually reflects retention and relearning/re-encoding. Another important point is that retention can only be examined after a word has been encoded. Thus, the ability to evaluate retention somewhat depends on encoding. If a child never encodes any words, retention cannot be measured. Likewise, if a child only encodes a few words (e.g., two), the measure of retention may not be as informative as the case where a child encodes a large number of words (e.g., 11).

### Successful Encoding and Retention

Encoding and retention profiles were explored for all 27 children at each naming test as treatment unfolded, and at the end of treatment based on cumulative performance throughout treatment. The group means and standard deviations were used to compute z-scores for each child for encoding and retention (i.e.,  $z = [\text{child score} - M] / SD$ ) at each test and at the end of treatment. The z-scores were then used to classify encoding and retention for each child at each test point as (a) low, based on a z-score at or below  $-1.00$ ; (b) average, based on a z-score between  $-1.00$  and  $+1.00$ ; or (c) high, based on a z-score at or above  $+1.00$ . Encoding and retention classification based on the cumulative treatment is shown in online Supplemental Material S1 (see Table S1). This overall classification of encoding and retention was then related to each child's overall treatment response to determine what aspect of learning limited treatment response. Individual classification is shown in online Supplemental Material S1 (see Table S2) and is summarized below in Table 5.

As shown in Table 5, the no treatment response group tended to exhibit low encoding, with 63% of the group characterized as low encoders. Retention was variable within this group, with relatively equal numbers of children within low versus average versus high retention. Turning to the average treatment response group, these children tended to exhibit average encoding, with 93% of the group characterized as average encoders. Likewise, the majority of the group showed average (50%) or high retention (29%), with few

**Table 5.** Mean, standard deviation, and range for number of words encoded and percentage of words retained, along with the percentage of children classified as having low, average, or high encoding or retention by treatment response group.

Measure	No treatment response	Average treatment response	High treatment response
<b>Encoding</b>			
<i>M</i>	4	8	15
<i>SD</i>	1	3	3
Range	2–6	4–15	12–19
% low	63	0	0
% average	38	93	20
% high	0	7	80
<b>Retention</b>			
<i>M</i>	72	79	89
<i>SD</i>	21	18	8
Range	50–100	50–100	79–100
% low	25	21	0
% average	38	50	80
% high	25	29	20

children exhibiting low retention (21%). Last, children in the high treatment response group were characterized as high encoders (80%) with average (80%) or high (20%) retention. Overall, the encoding profile appeared to be more closely related to the treatment response than the retention profile. Encoding may be particularly vulnerable when teaching new words via interactive book reading to children with SLI. Thus, modifications to the treatment to enhance encoding may be needed when treatment progress is inadequate. Memory evolution appeared to play a lesser role in treatment outcome, but it could be important for a small subset of children.

In addition to this examination of encoding and memory evolution based on the entire treatment, the same procedures were used to define low versus average versus high encoding and memory evolution at each naming test. These values are shown in Figure 4. Their use will be illustrated in the following section.

### ***Pulling It All Together: Clinical Decision Making***

The goal of this clinical forum was to provide guidance to clinicians on early benchmarks of successful word learning in an interactive book reading treatment and to examine encoding and retention during treatment in an attempt to determine how each contribute to treatment outcomes by kindergarten children with SLI. The results are summarized in Figure 4 as clinical decision-making trees. Overall, the following performance during our treatment suggests cause for concern and a potential need to modify the treatment: naming zero to one (out of 30) treated words correctly at Naming Test 1; naming zero to two (out of 30) treated words correctly at Naming Test 2; and naming zero to three (out of 30) treated words correctly at Naming Test 3. In addition, the results showed that more children in the same treatment response group had the same encoding classification than retention classification. This is consistent with findings from a prior study of adults with a history of

language impairment, where encoding was found to be the primary limiting factor in word learning, but memory evolution also contributed to word learning (McGregor, Licandro, et al., 2013). This finding also is consistent with the idea that memory evolution is somewhat dependent on encoding: A word cannot be retained if it hasn't first been encoded.

The remainder of the discussion applies the clinical decision-making framework, shown in Figure 4, to four specific children, shown in Figure 5, to illustrate how these (or similar) data can be used to inform clinical practice. Note that the data from all children are available in online Supplemental Material S1 (see Figures S1, S2, and S3). Figure 5 provides the same benchmarks for each child. The graph for each child shows the adequate progress minimum with a short-dashed line. These are the same cutoffs noted in Figure 4. The graph also displays performance of the children with an average or high treatment response using a long-dashed line. The solid dark line in each graph is the target child's performance. Likewise, the table shows specific words encoded or retained at each test point for a given child. All children illustrated in Figure 5 were in the 24 exposures condition. This intensity was chosen because there was high variability across children in treatment response, allowing illustration of a variety of treatment progress and encoding/memory evolution profiles.

Note that our benchmarks for each naming test are based on the unmodified version of the treatment that was administered as part of our research study. However, in this article, modifications are recommended for certain children. It is still appropriate to use benchmarks from the unmodified version of the treatment because the goal of the modifications is to tailor the treatment to a specific target child so that the target child achieves progress that is at least as good as the children who showed an average response. Stated differently, the goal of this entire process is to ensure that each child who is administered a treatment achieves at least the average treatment response and recognizes that some children will need different support to accomplish that.

### **First Naming Test**

The first naming test during treatment can only provide insights about encoding. To gain insights about memory evolution, a clinician needs evidence that a word was known at one naming test and then either known or not known at a subsequent naming test. Thus, data from the first naming test allow a clinician to evaluate how well a child is encoding the treated words as a result of the current treatment program but do not support evaluation of memory evolution. Any adjustments at this point in treatment will necessarily need to focus on encoding. This is illustrated in Figure 4. This also highlights one potential explanation of why encoding may play a stronger role than memory evolution in word learning during clinical treatment. To be specific, the child doesn't have an opportunity or doesn't have many opportunities to retain words if no or very few words are encoded. Thus, encoding is an important first step in getting word learning off the ground.

**Figure 4.** Clinical decision-making framework for performance at each early naming test.

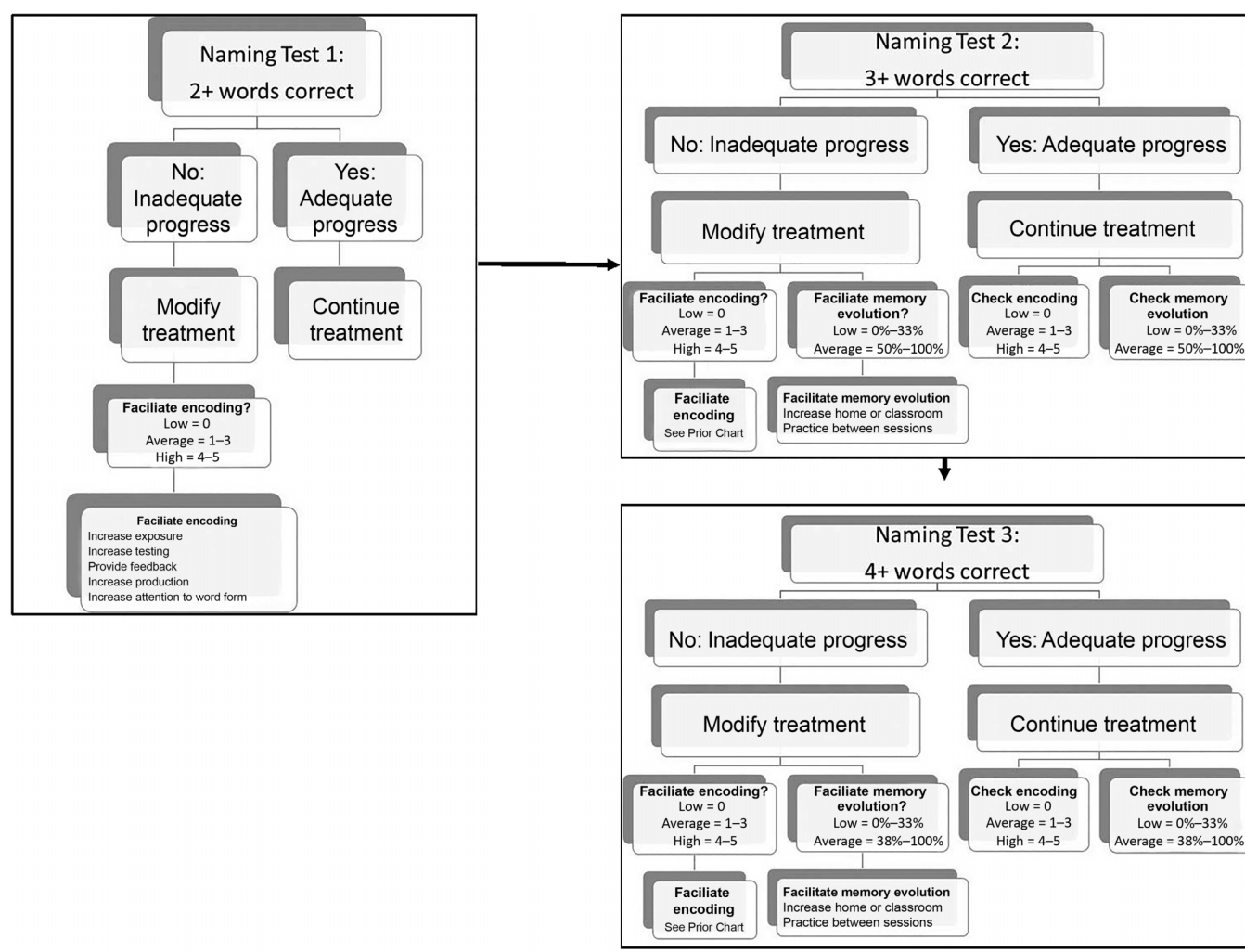
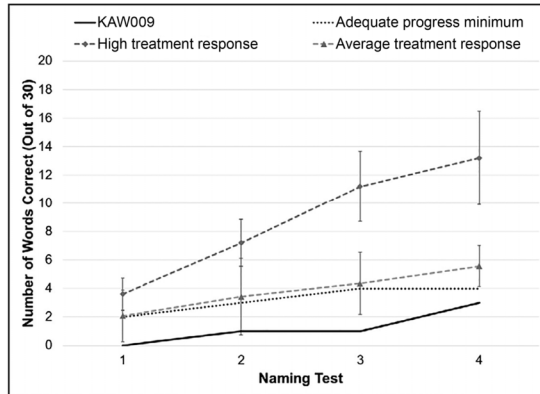


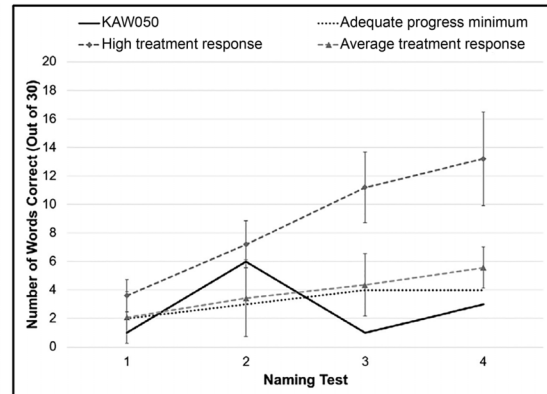
Figure 4 notes that adequate progress at the first naming test consists of naming two or more words correctly. Examining the children in Figure 5, KAW009 does not show adequate progress, naming zero words correctly at the first naming test. Likewise, KAW050 does not show adequate progress, naming one word correctly (i.e., *flashing*) at the first naming test. In contrast, KAW018 exceeds the criterion for adequate progress, naming four words (i.e., *invisible*, *pouted*, *snuggled*, and *squawked*) correctly at the first naming test. The graph illustrates that this performance is similar to that observed for children who ultimately have a high response to treatment. Likewise, Figure 4 notes that this performance constitutes high encoding. Last, KAW070 just meets the criterion for adequate progress, naming two words (i.e., *gloomy* and *swung*) correctly at the first naming test. The graph illustrates that this performance is similar to that observed for children who ultimately have an average response to treatment, and Figure 4 notes that this constitutes average encoding.

Taken together, two children (KAW009 and KAW050) are not making adequate progress. Based on this information, the clinician should likely modify the treatment for these children to promote better encoding. One possibility is that the number of exposures to the words can be increased. Prior experimental word learning studies with children with SLI suggest that their word learning can approximate that of a same-aged peer if given sufficient exposure (Gray, 2003; Rice et al., 1994). Likewise, our preliminary clinical trial showed that word learning improved as the number of exposures increased from 12 to 36 but then plateaued after 36 exposures (Storkel et al., 2017). These children were in the 24-exposure condition, thus an increase in exposure may have facilitated encoding. Even if a child were receiving the recommended 36 exposures, it is still possible that increasing the exposures would result in added benefit. Group data can provide general guidance on how to construct an effective treatment a priori, but often there is variability in treatment responding, requiring these types of online adjustments as

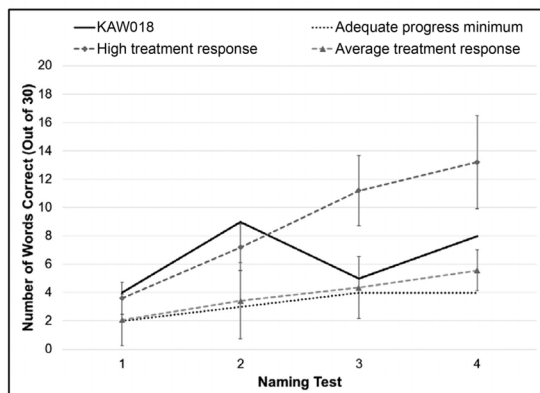
**Figure 5.** Data for four children in the 24-exposure condition: KAW009, KAW050, KAW018, and KAW070. Graphs in the top of each panel show the number of treated words named correctly by the target child (solid line), the average treatment response group (triangles with long-dashed line), and the high treatment response group (diamonds with long-dashed line). The minimum cutoff for adequate progress is shown by the short-dashed line. Tables in the bottom of each panel show the specific treated words that were named correctly for the first time (unshaded cells) or for a second or third time (shaded cells) at each naming test to illustrate encoding (unshaded cells) or memory evolution (shaded cells).



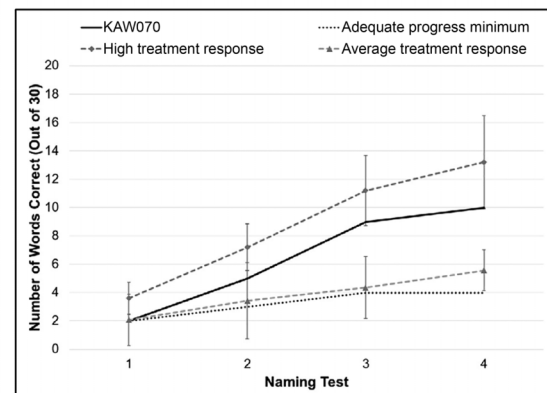
KAW009 Treated Words	1	2	3	4
Awful	0	1	1	1
Gathered	0	0	0	1
Haddock	0	0	0	1
Newly learned	0	1	0	2
Retained	N/A	N/A	1 of 1	1 of 1
% Retained			100%	100%



KAW050 Treated Words	1	2	3	4
Flashing	1	1	1	1
Awful	0	1	0	0
Crept	0	1	0	0
Snuggled	0	1	0	1
Squawked	0	1	0	1
Tight	0	1	0	0
Newly learned	1	5	0	0
Retained	N/A	1 of 1	1 of 6	3 of 6
% Retained		100%	17%	50%



KAW018 Treated Words	1	2	3	4
Invisible	1	1	0	1
Pouted	1	1	1	0
Snuggled	1	1	1	1
Squawked	1	1	0	1
Flashing	0	1	1	1
Haddock	0	1	0	1
Smooth	0	1	0	1
Tailor	0	1	0	1
Worm	0	1	0	0
Crept	0	0	1	0
Gulp	0	0	1	0
Tight	0	0	0	1
Newly learned	4	5	2	1
Retained	N/A	4 of 4	3 of 9	7 of 11
% Retained		100%	33%	64%



KAW070 Treated Words	1	2	3	4
Gloomy	1	1	1	1
Swung	1	1	0	0
Glared	0	1	1	1
Scarlet	0	1	1	1
Spotless	0	1	1	1
Furnace	0	0	1	1
Hooves	0	0	1	0
Marsh	0	0	1	1
Ripe	0	0	1	1
Sideline	0	0	1	1
Clamor	0	0	0	1
Overjoyed	0	0	0	1
Newly learned	2	3	5	2
Retained	N/A	2 of 2	4 of 5	8 of 10
% Retained		100%	80%	80%



the treatment unfolds for a particular child. Thus, if a clinician determined that a child was not making adequate progress with the current number of exposures, increasing the number of exposures could be beneficial. However, other options should be considered because our preliminary trial did show evidence of diminishing returns as the number of exposures increased beyond 36.

It is notable that our approach to interactive book reading was focused on input because it was modeled after a prior clinical trial that focused on input and because it focused heavily on determining the adequate number of exposures for children with SLI. A focus on input facilitated careful control of the number of exposures to the new words. However, this may not have been the best approach to support encoding. A more participatory approach may have facilitated encoding by children with SLI because this would have given the children practice retrieving the words from memory as well as feedback concerning incorrect responses (Butler, Karpicke, & Roediger, 2007; Grimaldi & Karpicke, 2012; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Thus, to enhance encoding a clinician might consider adding tests of meaning with feedback to the treatment. For example, the clinician could ask the child to provide a definition or synonym or to use the word to describe a picture. The clinician could then provide support if the child was unable to provide any response or could provide corrective feedback if the child's response contained inaccuracies.

Moreover, our instruction focused heavily on meaning, but there is evidence that children and adults with SLI may find word forms challenging (Gray, 2004; McGregor et al., 2016; McGregor, Licandro, et al., 2013). Therefore, a clinician might consider adding specific training and testing related to word forms. For example, the clinician could highlight the parts of the word (e.g., the onset, the rhyme) or phonological similarities and differences with known words (e.g., note that the target word rhymes with a common word that the child knows). The clinician also could ask the child to imitate the word as a means of practicing the word form. Last, the clinician could test any of this information (e.g., ask the child to identify the beginning sound of the word, ask the child to name a picture of the word) and provide support or corrective feedback.

### Second Naming Test

As shown in Figure 4, naming three or more words correctly at the second naming test constitutes adequate progress. Also shown in Figure 4, the second naming test offers an opportunity to evaluate both encoding and memory evolution, although the ability to evaluate memory evolution depends on how many words were named correctly at the first naming test—that is, if the child didn't name any words correctly at the first naming test, then there is no opportunity to evaluate memory evolution. This is the case for KAW009. For this child, only overall progress and encoding can be evaluated at the second naming test. In terms of overall progress, KAW009 named one word (i.e., *awful*) correctly at the second naming test. As noted

in Figure 4, this indicates inadequate progress. In terms of encoding, learning one new word just makes the cutoff for average encoding. However, given inadequate overall progress, it is likely that treatment should be modified to better support encoding, as previously detailed. Although inadequate progress was identified at the first naming test for KAW009 and modifications were suggested, no modifications were actually made for any of these children because they were part of a research protocol that could not be modified due to the objectives of the study. A clinician would likely further modify KAW009's treatment and may even want to consider a different treatment approach at this point given continued inadequate progress.

Turning to KAW050 in Figure 5, this child named six words (i.e., *flashing*, *awful*, *crept*, *snuggled*, *squawked*, and *tight*) correctly at the second naming test, demonstrating adequate progress. Four of these words were named correctly for the first time (i.e., *awful*, *crept*, *snuggled*, *squawked*, and *tight*), indicating high encoding (see Figure 4). Likewise, the one word that was named correctly at the first naming test (i.e., *flashing*) was named correctly again (i.e., 100% retention), indicating average memory evolution (see Figure 4). Evaluation of memory evolution is tentative because it is based on a very small sample (i.e., one word). Despite the lack of modifications to this child's treatment program, KAW050 demonstrated strong progress at the second naming test. Further modifications do not appear to be needed at this point.

Likewise, KAW018 continued to show adequate progress, naming nine words correctly (i.e., *invisible*, *pouted*, *snuggled*, *squawked*, *flashing*, *haddock*, *smooth*, *tailor*, and *worn*) at the second naming test. Five of the words were named correctly for the first time (i.e., *flashing*, *haddock*, *smooth*, *tailor*, and *worn*), indicating high encoding. In complement, the remaining four words had been named correctly at the first naming test (i.e., *invisible*, *pouted*, *snuggled*, and *squawked*), constituting 100% retention, which is average memory evolution. Thus, KAW018 continues to demonstrate strong progress, approximating that of the high treatment response group.

KAW070 also demonstrates continued adequate progress, naming five words correctly (i.e., *gloomy*, *swung*, *glared*, *scarlet*, and *spotless*). Three of these words are correct for the first time (i.e., *glared*, *scarlet*, and *spotless*), indicating average encoding, and two were previously named correctly (i.e., *gloomy* and *swung*), indicating average memory evolution. KAW070's performance is approximately in the middle of the average and high treatment response groups, suggesting that treatment should continue without modification.

### Third Naming Test

As shown in Figure 4, adequate progress at the third naming test is defined as naming four or more words correctly. In addition, the third naming test offers another opportunity to evaluate both encoding and retention, and by this point, all of the example children have named at least one word correctly. As shown in Figure 5, KAW009

named only one word correctly (i.e., *awful*) at the third naming test. This word was previously named correctly, indicating average memory evolution. However, no new words were named correctly for the first time, indicating continued problems with encoding. As shown in Figure 5, this child ultimately did not respond to the treatment, likely due to poor encoding. If modifications had been made to the treatment protocol early in treatment, it is possible that a better outcome may have been attained. However, if a pattern like this was observed even with modifications, the clinician would likely want to consider a different treatment approach.

KAW050 named only one word correctly (i.e., *flashing*) at the third naming test, indicating inadequate progress. Both encoding and memory evolution appeared to be problematic. No new words were named correctly for the first time, indicating low encoding, and only one of six words (17%) was named correctly for a second or third time, indicating low memory evolution. Thus, a clinician would likely consider modifications to facilitate encoding and memory evolution. In terms of encoding, the clinician now has conflicting information about KAW050's encoding. In particular, encoding was categorized as high at Naming Test 2 but is now categorized as low. One reason why encoding might be variable across the two naming tests may relate to the interaction between the words and encoding. That is, the first words encoded may have been easy in some way to encode, where easy could relate to a variety of stimulus factors (e.g., the context sentences invoked situations that were familiar to the child, the visuals provided strong support for encoding the meaning) or the state of the child's knowledge of those words at the start of treatment. To be specific, children's representations or knowledge of words is not all or none but rather represents a continuum of knowledge (McGregor, Friedman, Reilly, & Newman, 2002; McGregor, Newman, Reilly, & Capone, 2002). In this way, a child could have partial knowledge of a word that would not be sufficient to support correct naming of the word pretreatment but could support encoding during treatment even for a child who struggles with encoding. The remaining treatment words may now be harder than the initial words learned, either in terms of stimulus characteristics or in terms of the child's current knowledge of those words. These harder words may now stress the child's encoding.

There are multiple ways that memory evolution could be better supported during interactive book reading. However, it is first important to revisit the relationship between encoding and memory evolution. To be specific, there is evidence that strengths or weaknesses during encoding can affect memory evolution (McGregor et al., 2016; Storkel, 2015). That is, a memory that is strong at the end of a treatment session is more likely to be retained across a gap between training sessions (Storkel, 2015). Thus, a clinician may still want to consider enhancing both encoding and memory evolution even if difficulties with memory evolution appear to be the only issue. The previously described suggestions for enhancing encoding can guide this process. In

addition to enhancing encoding, the clinician may want to consider other adjustments to the treatment to directly affect memory evolution. There are two possible mechanisms to explain forgetting during memory evolution (see Storkel, 2015, for review). One explanation suggests that the memory is forgotten during the period between the end of the treatment session and the onset of sleep. Sleep is thought to be crucial for memory evolution. Thus, it is hypothesized that a memory needs to be strong and accurate prior to the onset of sleep so that the representation can be reinforced and enhanced during sleep. With this hypothesis in mind, a clinician might want to consider ways to enhance practice outside of the treatment session, especially daily home practice. Home practice would potentially allow the memory for the new word to be reactivated and enhanced prior to sleep. The clinician could send home hard copy or electronic picture cards of the target words and printed or audio exposures used to teach the words during treatment sessions so that the child could review the new words either interactively with the parent or more independently with electronic support (see Goldstein et al., 2016; Leacox & Jackson, 2012; Messier & Wood, 2015, for examples of technology-supported practice).

The second explanation of forgetting during memory evolution is that integration of the memory for the new word with the memories of known words during sleep leads to confusion between the new word and the known words (see Storkel, 2015, for review). This confusion can weaken the memory for the new word, leading to forgetting. Home or classroom practice after sleep has occurred could be useful because it would essentially reteach the word after integration with known words has occurred. The extra practice might support relearning of the new word and strengthening of the memory of the new word. In addition, the clinician could consider making explicit connections between the new word and known words during treatment sessions. This could help the child understand the similarities and differences between the new word and other words, facilitating successful integration of the new word into the child's lexicon.

In terms of the remaining two children in Figure 5, both show adequate progress. To be specific, KAW018 names five words correctly (i.e., *pouted*, *snuggled*, *flashing*, *crept*, and *gulp*). KAW018 names two words correctly for the first time (i.e., *crept* and *gulp*), which is average encoding, and names three of nine possible words (i.e., *pouted*, *snuggled*, and *flashing*) correctly for a second (or third) time (33% retention), indicating low memory evolution. To this point, KAW018 had demonstrated high encoding and average memory evolution. Although KAW018's progress at the third naming test is adequate, the graph in Figure 5 clearly shows a drop in performance. Why might this be? KAW018's errors for previously correct items may shed some light. For *squawked*, *haddock*, and *tailor*, KAW018 responded "I don't know" or "nothing," suggesting that retrieval of the word form may have been difficult. Prior suggestions to enhance retrieval through testing and feedback may be useful to increase success with these and other

words. KAW018's response for *worn* suggests some confusion between the target word and other similar words encountered during treatment. To be specific, KAW018 alternated between naming *worn* as *worn* or *torn*. The definition provided for *worn* during treatment was "to become thin or torn from use." Likewise, KAW018's response for *invisible* was a semantically related word, namely *disappeared*, although the erred word was not provided during training. Thus, for these two words, KAW018 may be experiencing semantic confusion during encoding and/or memory evolution. Explicit comparison of the similar words during training may facilitate encoding and memory evolution. Although KAW018 demonstrates adequate progress, a clinician may want to consider treatment modifications given the child's drop in performance at the third naming test. This case demonstrates the importance of continuously monitoring and adjusting treatment based on progress. New information may be revealed over the course of treatment, and clinicians need to take advantage of that to maximize treatment gains.

Turning to the last case, KAW070 names nine words correctly (i.e., *gloomy*, *glared*, *scarlet*, *spotless*, *furnace*, *hooves*, *marsh*, *ripe*, and *sidelines*), demonstrating adequate progress. In terms of encoding, five words were named correctly for the first time (i.e., *furnace*, *hooves*, *marsh*, *ripe*, and *sidelines*), which is categorized as high encoding. In addition, KAW070 names four words correctly for a second or third time (i.e., *gloomy*, *glared*, *scarlet*, and *spotless*), indicating average memory evolution (i.e., 80% retention). KAW070's progress is just within the lower end of the high treatment response group. Taken together, KAW070 continues to show strong progress, indicating that treatment modifications are not needed.

## Conclusion

The goal of this study was to establish benchmarks for adequate word learning progress by kindergarten children with SLI during our interactive book reading treatment. Our data suggest that the following progress is cause for concern and potential modification of treatment: naming zero to one (out of 30) treated words correctly at Naming Test 1; naming zero to two (out of 30) treated words correctly at Naming Test 2; and naming zero to three (out of 30) treated words correctly at Naming Test 3. Moreover, word learning profiles suggested that encoding played a greater role than memory evolution in word learning success by children with SLI, although memory evolution was an important factor for some children and should not be ignored. Cases were used to illustrate how this information can be applied in clinical settings to monitor word learning progress during this version of interactive book reading and to select appropriate treatment modifications. These illustrations demonstrate how a clinician's understanding of a child's strengths and weaknesses develop over the course of treatment, substantiating the importance of regular data collection and decision-making to ensure the best possible outcomes for a child.

## Acknowledgments

The project described was supported by National Institutes of Health Grants DC 012824, DC05803, and HD02528. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Health. We would like to thank staff of the Participant Recruitment and Management Core of the Center for Biobehavioral Neurosciences of Communication Disorders (DC05803) for assistance with recruitment of preschools and children; staff of the Research Design and Analysis Unit of the Lifespan Institute (supported by HD02528) for assistance with study design; and staff of the Word and Sound Learning Lab (supported by DC 012824) for their contributions to stimulus creation, participant recruitment, data collection, data processing, and reliability calculations.

## References

- Adi-Japha, E., & Abu-Asba, H. (2014). Learning, forgetting, and relearning: Skill learning in children with language impairment. *American Journal of Speech-Language Pathology*, 23, 696–707. [https://doi.org/10.1044/2014\\_AJSLP-13-0031](https://doi.org/10.1044/2014_AJSLP-13-0031)
- Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders*, 44, 173–185. <https://doi.org/10.1016/j.jcomdis.2010.09.00>
- American Speech-Language-Hearing Association. (1997). Guidelines for screening for hearing impairment: School-age children, 5 through 18 years. *Asha*, 4, IV-74ff–IV-74hh.
- American Speech-Language-Hearing Association. (n.d.). *Evidence-based practice*. Retrieved from <http://www.asha.org/Research/EBP/>
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273–281. <https://doi.org/10.1037/1076-898X.13.4.273>
- Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools*, 39, S110–S137.
- Crain-Thoreson, C., & Dale, P. S. (1999). Enhancing linguistic performance: Parents and teachers as book reading partners for children with language delays. *Topics in Early Childhood Special Education*, 19, 28–39.
- Crowe, L. K., Norris, J. A., & Hoffman, P. R. (2004). Training caregivers to facilitate communicative participation of preschool children with language impairment during storybook reading. *Journal of Communication Disorders*, 37, 177–196. <https://doi.org/10.1016/j.jcomdis.2003.09.001>
- Dale, P. S., Crain-Thoreson, C., Notari-Syverson, A., & Cole, K. (1996). Parent-child book reading as an intervention technique for young children with language delays. *Topics in Early Childhood Special Education*, 16, 213–235.
- Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21, 803–820. <https://doi.org/10.1162/jocn.2009.21059>
- Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London: Series B, Biological Science*, 364, 3773–3800. <https://doi.org/10.1098/rstb.2009.0111>
- Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89, 105–132.



- Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a pre-school classroom. *Journal of Speech and Hearing Research*, 37, 913–923.
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation—Second Edition*. San Antonio, TX: Pearson.
- Goldstein, H., Kelly, E., Greenwood, C., McCune, L., Carta, J., Atwater, J., ... Spencer, T. (2016). Embedded instruction improves vocabulary learning during automated storybook reading among high-risk preschoolers. *Journal of Speech, Language, and Hearing Research*, 59, 484–500. [https://doi.org/10.1044/2015\\_JSLHR-L-15-0227](https://doi.org/10.1044/2015_JSLHR-L-15-0227)
- Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research*, 46, 56–67. [https://doi.org/10.1044/1092-4388\(2003/005\)](https://doi.org/10.1044/1092-4388(2003/005))
- Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research*, 47, 1117–1132. [https://doi.org/10.1044/1092-4388\(2004/083\)](https://doi.org/10.1044/1092-4388(2004/083))
- Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, 40, 505–513. <https://doi.org/10.3758/s13421-011-0174-0>
- Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology*, 116, 572–592. <https://doi.org/10.1016/j.jecp.2013.07.004>
- Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools*, 36, 17–32.
- Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53, 739–756.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968. <https://doi.org/10.1126/science.1152408>
- Komesidou, R., & Storkel, H. L. (2015). Learning and remembering new words: Clinical illustrations from children with specific language impairment. *Perspectives on Language Learning and Education*, 22, 138–146.
- Leacock, L., & Jackson, C. W. (2012). Spanish vocabulary-bridging technology-enhanced instruction for young English language learners' word learning. *Journal of Early Childhood Literacy*, 14, 175–197. <https://doi.org/10.1177/1468798412458518>
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80, 300–335. <https://doi.org/10.3102/0034654310377087>
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457. <https://doi.org/10.1037/0033-295x.102.3.419>
- McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2016). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*. Advance online publication. <https://doi.org/10.3109/17549507.2016.1159337>
- McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic representation and naming in young children. *Journal of Speech, Language, and Hearing Research*, 45, 332–346.
- McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, 56, 1845–1856. [https://doi.org/10.1044/1092-4388\(2013/12-0233\)](https://doi.org/10.1044/1092-4388(2013/12-0233))
- McGregor, K. K., Newman, R. M., Reilly, R., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 45, 998–1014.
- McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders*, 48, 307–319. <https://doi.org/10.1111/1460-6984.12008>
- Messier, J., & Wood, C. (2015). Facilitating vocabulary acquisition of children with cochlear implants using electronic storybooks. *Journal of Deaf Studies and Deaf Education*, 20, 356–373. <https://doi.org/10.1093/deafed/env031>
- Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, 79, 979–1007.
- Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education and Development*, 19, 7–26.
- Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, 44, 472–488. <https://doi.org/10.1177/0022219411414010>
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110, 611–646. <https://doi.org/10.1037/0033-295x.110.4.611>
- O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, 10, 389–397. [https://doi.org/10.1002/1098-1063\(2000\)10:4<389::aid-hipo5>3.0.co;2-p](https://doi.org/10.1002/1098-1063(2000)10:4<389::aid-hipo5>3.0.co;2-p)
- Oetting, J. B. (1999). Children with SLI use argument structure cues to learn verbs. *Journal of Speech, Language, and Hearing Research*, 42, 1261–1274.
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554–566.
- Pile, E. J. S., Girolametto, L., Johnson, C. J., Chen, X., & Cleave, P. J. (2010). Shared book reading intervention for children with language impairment: Using parents-as-aides in language intervention. *Canadian Journal of Speech-Language Pathology and Audiology*, 34, 96–109.
- Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales and the Reynolds Intellectual Screening Test*. Lutz, FL: Psychological Assessment Resources.
- Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, 37, 106–122.
- Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research*, 48, 1397–1411.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention.



- Psychological Science*, 170, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *Clinical Evaluation of Language Fundamentals—Fourth Edition*. San Antonio, TX: Pearson.
- Seymour, H. N., Roeper, T. W., de Villers, J., & de Villers, P. A.** (2005). *Diagnostic Evaluation of Language Variation—Norm-Referenced*. San Antonio, TX: Pearson.
- Storkel, H. L.** (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321–1337.
- Storkel, H. L.** (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research*, 46, 1312–1323.
- Storkel, H. L.** (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology*, 17, 1–12.
- Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y.** (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, 56, 1689–1700.
- Storkel, H. L., & Lee, S.-Y.** (2011). The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, 26, 191–211.
- Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S.** (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in School*, 48, 16–30. [https://doi.org/10.1044/2016\\_LSHSS-16-0014](https://doi.org/10.1044/2016_LSHSS-16-0014)
- Tamminen, J., & Gaskell, M. G.** (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology*, 66, 1001–1025. <https://doi.org/10.1080/17470218.2012.724694>
- Vlach, H. A., & Sandhofer, C. M.** (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, 3, 46. <https://doi.org/10.3389/fpsyg.2012.00046>
- Voelmle, K., & Storkel, H. L.** (2015). Teaching new words to children with specific language impairment using interactive book reading. *Perspectives on Language Learning and Education*, 22, 131–137.
- Wagner, R. K., Torgesen, J., Rashotte, C., & Pearson, N. A.** (2013). *Comprehensive Test of Phonological Processing—Second Edition*. Austin, TX: Pro-Ed.
- Walker, M. P., & Stickgold, R.** (2010). Overnight alchemy: Sleep-dependent memory evolution. *Nature Reviews Neuroscience*, 11, 218. <https://doi.org/10.1038/nrn2762-cl>
- Wallace, G., & Hammill, D. D.** (2013). *Comprehensive Receptive and Expressive Vocabulary Test—Third Edition*. Austin, TX: Pro-Ed.
- Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, M. C., Valdez-Menchaca, M. C., & Caulfield, M.** (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24, 552–559.
- Whitehurst, G. J., Fischel, J. E., Lonigan, C. J., Valdez-Menchaca, M. C., Arnold, D. S., & Smith, M.** (1991). Treatment of early expressive language delay: If, when, and how. *Topics in Language Disorders*, 11, 55–68.